Technical University of Denmark

A multi-dimensional dynamic linear model for monitoring slaughter pig production

Jensen, Dan Børge; Cornou, Cecile; Toft, Nils; Kristensen, Anders

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Jensen, D. B., Cornou, C., Toft, N., & Kristensen, A. (2015). A multi-dimensional dynamic linear model for monitoring slaughter pig production. Paper presented at 7th European Conference on Precision Livestock Farming (EC-PLF), Milan, Italy.

A multi-dimensional dynamic linear model for monitoring slaughter pig production

Conference Paper · September 2015

4 authors:

Dan B Jensen
University of Copenhagen

Cecile Cornou
Novo Nordisk

Nils Toft
Technical University of Denmark

Anders Ringgaard Kristensen
University of Copenhagen

A multi-dimensional dynamic linear model for monitoring slaughter pig production

Dan Børge Jensen[1,*], Cécile Cornou[1], Nils Toft[2], Anders Ringgaard Kristensen[1]
[1]*Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 2, DK-1870 Frederiksberg C, Denmark*
[2]*National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark*
daj@sund.ku.dk

Abstract

Scientists and farmers still lack an efficient way to unify the large number of different types of data series, which are increasingly being generated in relation to automatic herd monitoring. Such a unifying model should be able to account for the correlations between the various types of data, resulting in a model which could potentially yield more information than can be gained from the individual components separately. Here we present such a model for monitoring slaughter pig production, in the form of a multivariate dynamic linear model. This model unifies three types of data (live weight, feed- and water consumption), measured at different levels of detail (individual pig and double-pen level) and with different observational frequencies (weekly and daily), using series collected for the Danish PigIT project. The presented three-dimensional model serves as a proof of concept, and it should be straightforward to expand it with additional data types.

Key words

Dynamic linear model, information unification, modeling, monitoring, pig production

Introduction

For many years, a whole range of sensors have been available for monitoring variables relevant for *e.g.* mastitis detection in dairy cows (Viguier *et al.* 2009) and the application of sensor technology is slowly being introduced for pig herd monitoring. The idea is that the collected data, combined with a proper alarm system, can provide the farmer with early alarms, thus allowing proper proactive responses to undesired changes in the herd. However, there are serious issues with the methods currently described. First, the trade-offs of sensitivity and specificity are generally unacceptable (Hogeveen *et al.* 2010). Furthermore, the systems tend to consider each monitored variable in isolation, so that an alarm is based on the value of just one variable, and no interaction effects are considered. We therefore believe that better integration of the available information could yield better methods for prediction of animal health states.

We suggest employing a multivariate dynamic linear model (DLM) (West & Harrison 1997) as a means of obtaining a more holistic monitoring of animal herds, taking into account the interconnectedness of all the variables of interest. Univariate DLM's have previously been attempted for automated estrus detection in sows (Ostersen *et al.* 2010) and a multivariate DLM has been used to predict litter sizes in sows (Bono *et al.* 2012), but where only one type of information was considered. In general the use of DLM is under-utilized in the fields of animal- and veterinary science.

This paper serves as a proof of concept, demonstrating the use of a multivariate DLM for monitoring slaughter pigs in a Danish finisher unit, in terms of live weight, feed usage and water usage.

Materials and Methods

Data source

The work described in this paper was done using data collected for the PigIT Project[1] in a commercial Danish pig farm. Specifically, the data were collected in the farm's finisher unit, housing slaughter pigs while they grow from roughly 30-100 kg. The unit consists of five sections, each with 14 pens. Each pen contains 18 pigs (at insertion), sorted by sex and size. The climate within each section is controlled by a combi-diffuse ventilation system, computer-controlled sprinklers above each pen and heating pipes installed in the back walls.
For the PigIT Project, a number of sensors have been installed to automatically record data on feed usage, water flow to the drinking nipples and temperature in 16 of the 70 pens in the finisher unit.

Liquid feed is dispensed automatically into troughs, shared between two neighboring pens, as seen on Figure 1 A. Two such feed-sharing pens will be referred to as a *double pen*. The expected amount of liquid feed required for a given double pen is adjusted regularly by manual observation of how much of the dispensed feed has been left uneaten.
Water is dispensed from drinking nipples which, like the feed dispensers, are also shared between two pens in a double pen, as seen in Figure 1 B. In the 16 PigIT-pens (eight double pens) flow meters are installed above the water nipple to measure water flow to the double pen.

In addition, the individual pigs from two double pens in the same section (section 2) are manually weighed once per week from insertion until the first pigs from that section are sent to the abattoir. The weight measurements are performed with the pig scale depicted on Figure 1 C. The pigs in these pens are individually identified with RFID ear tags.

The data applied in this paper were collected from those two double pens, where live weight was recorded in addition to feed consumption and water flow. The used data were collected between November 20th 2013 and December 12th 2014, and included four separate insertions of new pigs. Thus the following models were based on a total of eight separate sets of observations.

A) B) C)

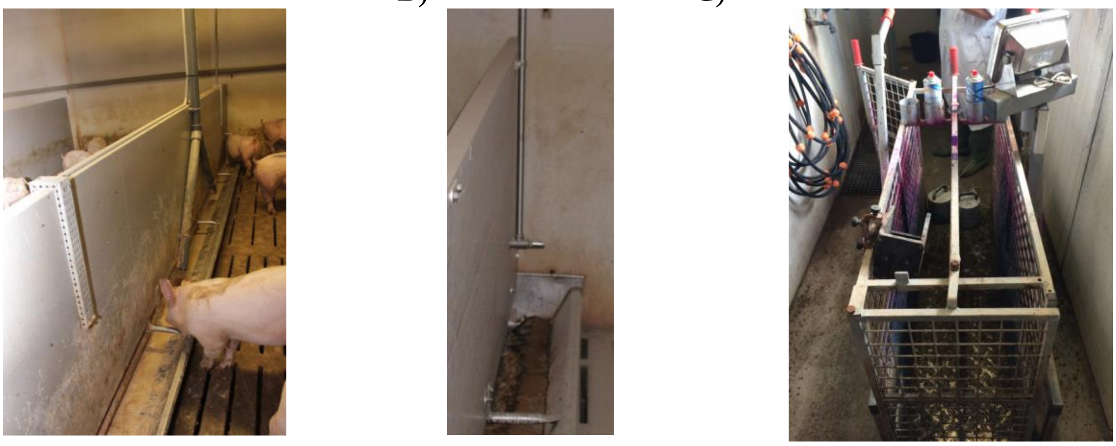

Figure 1: The sources of the data used in this paper. A) Liquid feed dispensed to the double pen by the feeding system. B) Water consumed by the pigs in a double pen, recorded by a flow meter above drinking nipple. C) Scale for manual recording of individual animal weights.

[1] http://pigit.ku.dk/

Modeling

All analysis, modeling and representations of the data were done using the software R, a language and environment for statistical computing (The R Core Team 2013).

In this paper, we aim to demonstrate a method for meaningfully combining multiple different kinds of observational data (live weight, feed usage and water flow). From farm records on insertion and removal of individual pigs, it was possible to know how many pigs were in a given pen at any given time, and this information was used to normalize the feed usage and water flow to daily averages per pig in the double pen. Although the live weights of the pigs were recorded individually, these too were aggregated to a per pig average for the double pen, in order to simplify the model.

A multivariate dynamic linear model (DLM), as described by West and Harrison (1997), was the method chosen to combine the data. In general, a DLM consists of an observation equation and a system equation (Equations (1) and (2), respectively).

$$Y_t = F_t'\theta_t + v_t, \qquad v_t \sim N\left(\underline{0}, V\right) \tag{1}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \qquad w_t \sim N\left(\underline{0}, W\right) \tag{2}$$

Equation (1) describes how the values of an observation vector (Y_t) depends on an unobservable parameter vector (θ_t) to time t. The unit of time used in this model was one day.

In our case, the parameter vector contains the estimated underlying values for live weight (*LW*), feed usage (*Feed*) and water flow (*Water*), as well as the rates at which those same values change at time t (*dLW, dFeed, dWater*, respectively), as seen in Table 1, θ_t. The underlying values at time t were estimated using a Kalman filter as described by West and Harrison (1997). In short, the Kalman filter is a method for filtering noise from the data by considering the actual observations, the error in the model forecasts and the systematic and observational variances.

For our purpose, the (transposed) design matrix has a structure with a basis as seen in Table 1 (F_t'). This structure serves to separate the estimated underlying values of the live weight, feed usage and water flow in the parameter vector from their respective trend values, in accordance with Equation (1). The structure is varied according to which variables are observed for a given time t, with the first, second and third row being included when live weight, feed usage and water flow are observed, respectively. Thus missing observations will be ignored when the parameter vector is updated.

The structure of the system matrix (G_t) is constant in our case (Table 1, G_t). This structure serves to add the trend values to the corresponding estimated values of the three parameters of interest, thus updating their values from time $t-1$ to time t, in accordance with Equation (2).

The initial values of live weight, feed usage and water flow were estimated from all available data as the average, normalized values observed on the first day of a batch insertion. The initial growth trend for live weight and feed usage were estimated as the average daily change in those values between the first and eighth day of observation. The water flow was seen to vary greatly from day to day, but did not follow any general trend over the grower/finisher periods. The initial rate of change was therefore set to 0.

Table 1: The structures of the three matrices, presented in Equations (1) and (2), as they apply to the data used for this paper.
$\boldsymbol{\theta_t}$: The Parameter Vector. $\boldsymbol{F_t'}$: The Design Matrix (transposed). $\boldsymbol{G_t}$: The System Matrix

θ_t	F_t'	G_t
$\begin{bmatrix} LW_t \\ dLW_t \\ Feed_t \\ dFeed_t \\ Water_t \\ dWater_t \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

The observational co-variance matrix (V) and the systematic co-variance matrix (W) were estimated from all available data, using the expectation maximization (EM) algorithm, as described by West and Harrison (1997), until convergence, which by visual inspection was found to occur after 50,000 iterations.

Unification of model forecast errors

Once the parameters defining the DLM had been estimated, the average live weight, feed usage and water flow per pig in each of the eight separate batch observation series were modeled. During the modeling, a vector (e_t) of forecast errors (*Observed values – Forecasted values*) was continuously generated for each time step. In addition, a matrix describing the forecast co-variances (Q_t) was continuously generated, as described by West and Harrison (1997). A Cholesky decomposition was calculated for Q_t using the R function `chol`. Using the decomposed matrix (C_{Qt}), the error vector was transformed, as seen in Equation (3).

$$u_t = C_{Qt}^{-1} \cdot e_t \tag{3}$$

This transformation ensures that the transformed error values in u_t are mutually independent and each follow a standard normal distribution. Thus a single value measuring the square of deviation from 0, the mean within this frame of reference, can be easily calculated for the set of forecast errors, as seen in Equation (4).

$$d_t^2 = u_t' \cdot u_t \tag{4}$$

This unified error will follow a $\mathcal{X}^2$ distribution with n degrees of freedom, where n is the number of elements in u_t. Thus d_t^2 can be plotted to a conventional Shewhart control chart (Montgomery 2005) to allow for an easy monitoring of the complex system. The upper control limit was set to the 0.99 quantile of the $\mathcal{X}^2$ distribution. To allow for a constant control limit in response to varying degrees of freedom, d_t^2 was adjusted, according to Equation (5).

$$d_{adj.t}^2 = d_t^2 \cdot \left(\frac{\mathcal{X}^2(0.99, 3)}{\mathcal{X}^2(0.99, n)} \right) \tag{5}$$

Results and discussion

Model parameter values

The estimated initial values of the parameter vector are seen in Equation (6).

$$\theta_0 = (29.0\,, 0.65\,, 3.3\,, 0.79\,, 0.6\,, 0.0)' \tag{6}$$

As is seen, the average pig initially weigh 29 kg, grow at a rate of 650 grams per day and eats 3.3 kg feed per day with a daily increase of 790 grams. The normalized water flow to the double pen is 0.6 liters per day per pig, with no (0.0) systematic daily change.

The matrices describing the observational and systematic co-variances are seen in Equations (7) and (8), respectively.

$$V = \begin{bmatrix} 2.37 & 5.22 & -3.49e-4 \\ & 10.43 & -2.23e-4 \\ & & 6.62e-5 \end{bmatrix} \tag{7}$$

$$W = \begin{bmatrix} 1.11 & 1.70e-3 & -1.41 & -2.80e-3 & 1.48e-2 & 7.37e-6 \\ & 2.61e-6 & -5.00e-4 & -3.53e-6 & -1.88e-5 & 0.00 \\ & & 4.52 & 1.37e-2 & -5.30e-2 & -3.31e-5 \\ & & & 5.92e-5 & -1.00e-4 & 2.53e-9 \\ & & & & 1.61e-1 & 2.17e-3 \\ & & & & & 1.61e-8 \end{bmatrix} \tag{8}$$

Notice that V has a 3x3 structure, consistent with the three values which can be observed at each time t, while W has a 6x6 structure, consistent with the six values in the parameter vector.
It is worth noting that the diagonal values in both matrices would have been the same if each variable of interest had been modeled separately. It is thus the co-variances outside the diagonals which provide the extra information about the interconnectedness of each of the monitored variables.

Modeling

The DLM defined as described in the previous sections was used to model each of the eight available sets of batch data. Figure 2 shows three notable examples of the output of this modeling. These are the batches inserted on July 7th 2014 to pen number 2.5 and 2.10 (top and bottom row, respectively) and the batch inserted on October 9th 2014 to pen number 2.10 (middle row).

The left column of Figure 2 shows the observed values for mean live weight (circles), feed usage (triangles) and water flow (solid squares). In addition, the left column shows the filtered mean, as estimated by the DLM, for live weight (solid line), feed usage (thick dashed line) and water flow (dotted line).

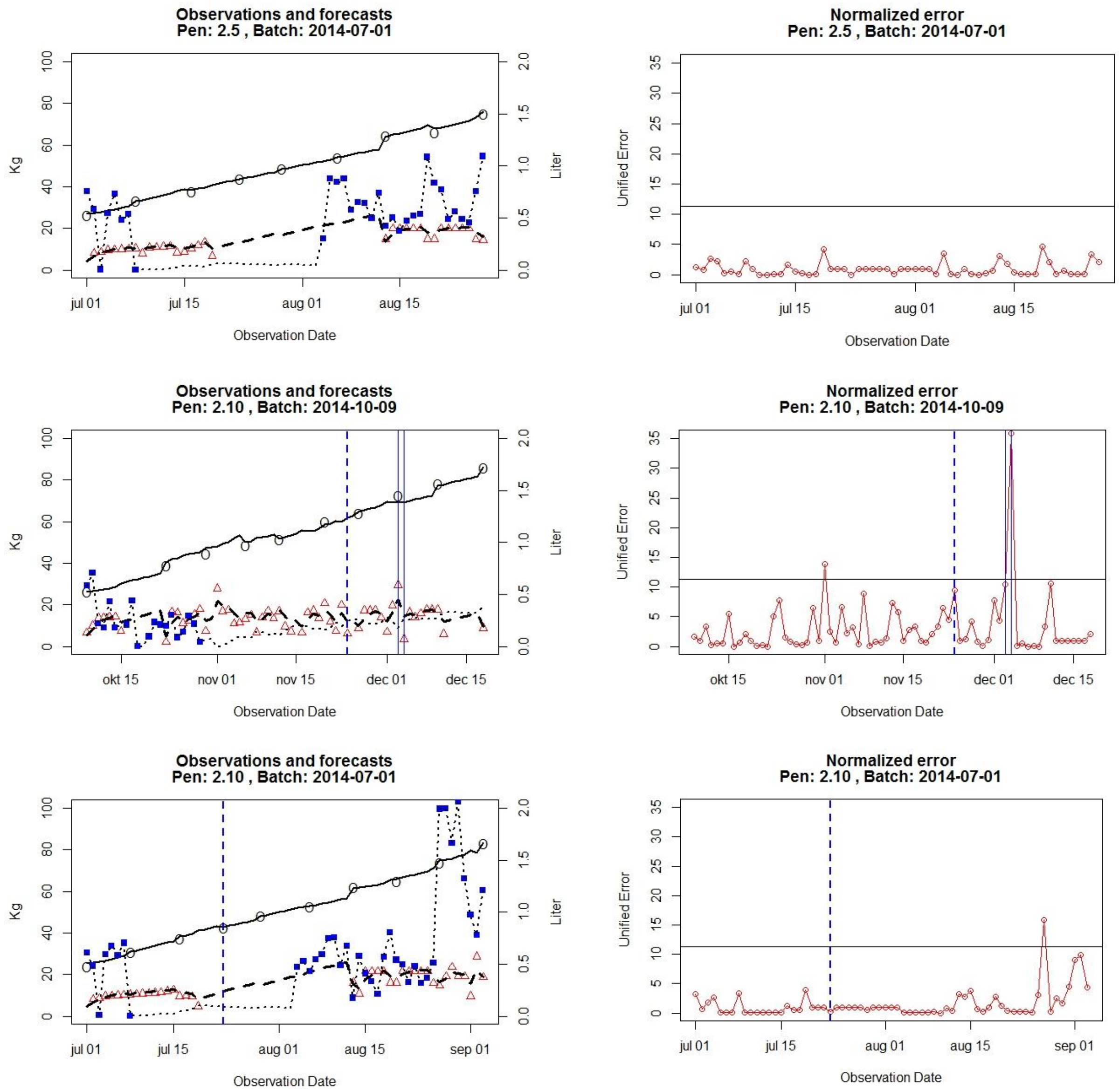

Figure 2: Left column: the observed values of mean live weight (circles), feed usage (triangles) and water flow (solid squares) per pig in three separate batches. In addition, the filtered mean values, estimated by the model, for live weight (solid line), feed usage (dashed line) and water flow (dotted line). Right column: the unified forecast error for mean live weight, feed usage and water flow per pig, corresponding to the observations depicted in the left column. Vertical lines in both columns: observations of diarrhea (dashed) and pen fouling (solid).

The right column shows the Shewhart control charts of the adjusted unified forecasts errors (circles connected by red lines), according to Equation (5). The horizontal lines in the control charts show the control value, *i.e.* the 0.99 quantile of the $\mathcal{X}^2$ distribution with 3 degrees of freedom (11.34).
For both columns, observations of diarrhea and pen fouling are marked by vertical lines. Diarrhea is marked by thick dashed lines, while solid lines represent pen fouling.

As is seen, the output in the top row is from a batch where no undesired events were observed, and all unified errors are all well below the control limit.

For the middle batch, pen fouling is observed twice (on the 3rd and 4th of December) and one case of diarrhea is observed around the 20th of November. The first pen fouling event falls just below the control line, but the second one coincides with a very clear spike in the unified error, which would yield a successful alarm. However, the system fails to raise an alarm about the diarrhea. This is probably because this event occurs during a period of time where water data is not available, which would be expected to strongly correlate with diarrhea.

The bottom batch provides two interesting examples of how the system can fail in its function. First, an undesired event (diarrhea) is observed at July 20th. This event happens to coincide with a relatively long period of time where the data on feed usage and water flow are both missing, and only the weight observation of that day is available. It can only be assumed that this information, especially regarding water flow, would have contributed to a more extreme unified error, and thus this example illustrates the value of having functioning sensors throughout a monitoring period.
Conversely, around the 25th of August, a tall peak is seen in the unified error, in spite of there being no observed undesired events. This peak is seen to be caused by a sudden dramatic and uncharacteristic increase in the water flow. From temperature records it can be found that this increase in water flow coincides with a sudden increase in temperature inside the double pen 2.10, while a similar temperature increase was not experienced in double pen 2.5 (data not shown).

Perspectives

When employing a dynamic linear model, an uncharacteristically large forecast error (here the limit was set to 11.34), is an indication that the observed system has changed significantly from the assumptions of the model. Thus, if the model has been optimized for describing a perfectly healthy batch of pigs, uncharacteristic forecast errors would likely indicate an outbreak of disease. To what extend the method demonstrated in this paper allow for more accurate disease detection, compared to other methods, will be a subject for further studies.

It should be noted that the form of forecast error unification demonstrated here can only yield an absolute magnitude of the error, and thus, unlike conventional univariate control charts, this control system cannot take into account whether some errors are positive and others negative. This potential problem could be circumvented by parsing the separate, non-unified errors to other classification systems, *e.g.* artificial neural networks or Bayesian classifiers. This would require separate training and validation of these systems, in addition to what is needed for the DLM itself. However, it is conceivable that such parsing could yield better detection of undesired events, and even allow for specific error patterns to be mapped to specific conditions, which would be another natural objective for further studies.

Whether or not a multivariate DLM defined from one herd can be directly applied in another herd or between different breeds of pigs, and how often such models need to be updated to keep up with the biological changes from breeding, are additional questions requiring further studies to answer.

Furthermore, here we have demonstrated the method with three measurable variables, but it would be trivial to adapt the model to include more (or fewer) lines of evidence, depending on data availability. We could envision modeling the live weight of each pig in the double pen individually or including the modeling of some measure of activity captured by video, etc. All that is needed is to design the appropriate design- and system matrices and the availability of the relevant data.

Lastly, this paper showcased the use of a multivariate DLM for monitoring slaughter pig production, but this method could just as well be employed in any animal production where data is routinely collected. An obvious example is dairy production, where several lines of data are often collected while milking the cows, but where a good standard for combining this data for meaningful information extraction is still lacking (Rutten *et al.* 2013).

Conclusions

We show that one can meaningfully co-model three very different types of monitoring data (live weight, feed usage and water flow) from an animal production herd, using a multivariate dynamic linear model. The errors in the forecasts produced by such a model can be unified to allow for easy monitoring of the health state of the herd using a Shewhart control chart to raise appropriate alarms.

Acknowledgements

This research was carried out with support from The Danish Council for Strategic Research (The PigIT Project, Grant number 11-116191)

References

Bono, C., Cornou, C. & Kristensen, A.R., 2012. Dynamic production monitoring in pig herds I: Modeling and monitoring litter size at herd and sow level. *Livestock Science*, 149(3), pp.289–300.

Hogeveen, H., Kamphuis, C., Steeneveld, W. & Mollenhorst, H., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009.

Montgomery, D., 2005. *Introduction to Statistical Quality Control* 5th ed., Hoboken, NJ, USA: Wiley.

Ostersen, T., Cornou, C. & Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 74(1), pp.51–58.

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W. & Hogeveen, H., 2013. Invited review: sensors to support health management on dairy farms. *Journal of dairy science*, 96(4), pp.1928–52.

The R Core Team, 2013. *R : A Language and Environment for Statistical Computing* 2.15.3 ed., R Foundation for Statistical Computing.

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K. & O'Kennedy, R., 2009. Mastitis detection: current trends and future perspectives. *Trends in biotechnology*, 27(8), pp.486–93.

West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.